OIPE

RAW SEQUENCE LISTING   DATE: 05/13/2002
PATENT APPLICATION: US/10/023,066A   TIME: 09:31:20

Input Set : **A:\BB1037 US DIV CORRECTED SEQ LISTING.txt**
Output Set: **N:\CRF3\05102002\J023066A.raw**

SEQUENCE LISTING

```
 3 (1) GENERAL INFORMATION:
 5       (i) APPLICANT: E. I. DU PONT DE NEMOURS AND
 6                         COMPANY
 8      (ii) TITLE OF INVENTION: CHIMERIC GENES AND METHODS FOR
 9                              INCREASING THE LYSINE AND
10                              THREONINE CONTENT OF THE SEEDS OF
11                              PLANTS
13     (iii) NUMBER OF SEQUENCES: 107
15      (iv) CORRESPONDENCE ADDRESS:
16           (A) ADDRESSEE: E. I. DU PONT DE NEMOURS
17                            AND COMPANY
18           (B) STREET: 1007 MARKET STREET
19           (C) CITY: WILMINGTON
20           (D) STATE: DELAWARE
21           (E) COUNTRY: U.S.A.
22           (F) ZIP: 19898
24       (v) COMPUTER READABLE FORM:
25           (A) MEDIUM TYPE: FLOPPY DISK
26           (B) COMPUTER: IBM PC COMPATIBLE
27           (C) OPERATING SYSTEM: PC-DOS/MS-DOS
28           (D) SOFTWARE: MICROSOFT WORD VERSION 2.0C
30      (vi) CURRENT APPLICATION DATA:
```
C--> 31   **(A) APPLICATION NUMBER: US/10/023,066A**
C--> 32   **(B) FILING DATE: 29-Apr-2002**
```
33           (C) CLASSIFICATION:
35    (viii) ATTORNEY/AGENT INFORMATION:
36           (A) NAME: BARBARA C. SIEGELL
37           (B) REGISTRATION NUMBER: 30,684
38           (C) REFERENCE/DOCKET NUMBER: BB-1037-C
40      (ix) TELECOMMUNICATION INFORMATION:
41           (A) TELEPHONE: 302-992-4931
42           (B) TELEFAX: 302-773-0164
43           (C) TELEX: 835420
45 (2) INFORMATION FOR SEQ ID NO: 1:
47       (i) SEQUENCE CHARACTERISTICS:
48           (A) LENGTH: 1350 base pairs
49           (B) TYPE: nucleic acid
50           (C) STRANDEDNESS: single
51           (D) TOPOLOGY: linear
53      (ii) MOLECULE TYPE: DNA (genomic)
55      (ix) FEATURE:
56           (A) NAME/KEY: CDS
```

Input Set : A:\BB1037 US DIV CORRECTED SEQ LISTING.txt
Output Set: N:\CRF3\05102002\J023066A.raw

```
57          (B) LOCATION: 1..1350
59      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
61 ATG GCT GAA ATT GTT GTC TCC AAA TTT GGC GGT ACC AGC GTA GCT GAT    48
62 Met Ala Glu Ile Val Val Ser Lys Phe Gly Gly Thr Ser Val Ala Asp
63    1               5                  10                  15
65 TTT GAC GCC ATG AAC CGC AGC GCT GAT ATT GTG CTT TCT GAT GCC AAC    96
66 Phe Asp Ala Met Asn Arg Ser Ala Asp Ile Val Leu Ser Asp Ala Asn .
67              20                  25                  30
69 GTG CGT TTA GTT GTC CTC TCG GCT TCT GCT GGT ATC ACT AAT CTG CTG   144
70 Val Arg Leu Val Val Leu Ser Ala Ser Ala Gly Ile Thr Asn Leu Leu
71          35                  40                  45
73 GTC GCT TTA GCT GAA GGA CTG GAA CCT GGC GAG CGA TTC GAA AAA CTC   192
74 Val Ala Leu Ala Glu Gly Leu Glu Pro Gly Glu Arg Phe Glu Lys Leu
75       50                  55                  60
77 GAC GCT ATC CGC AAC ATC CAG TTT GCC ATT CTG GAA CGT CTG CGT TAC   240
78 Asp Ala Ile Arg Asn Ile Gln Phe Ala Ile Leu Glu Arg Leu Arg Tyr
79 65                  70                  75                  80
81 CCG AAC GTT ATC CGT GAA GAG ATT GAA CGT CTG CTG GAG AAC ATT ACT   288
82 Pro Asn Val Ile Arg Glu Glu Ile Glu Arg Leu Leu Glu Asn Ile Thr
83                  85                  90                  95
85 GTT CTG GCA GAA GCG GCG GCG CTG GCA ACG TCT CCG GCG CTG ACA GAT   336
86 Val Leu Ala Glu Ala Ala Ala Leu Ala Thr Ser Pro Ala Leu Thr Asp
87              100                 105                 110
89 GAG CTG GTC AGC CAC GGC GAG CTG ATG TCG ACC CTG CTG TTT GTT GAG   384
90 Glu Leu Val Ser His Gly Glu Leu Met Ser Thr Leu Leu Phe Val Glu
91          115                 120                 125
93 ATC CTG CGC GAA CGC GAT GTT CAG GCA CAG TGG TTT GAT GTA CGT AAA   432
94 Ile Leu Arg Glu Arg Asp Val Gln Ala Gln Trp Phe Asp Val Arg Lys
95       130                 135                 140
97 GTG ATG CGT ACC AAC GAC CGA TTT GGT CGT GCA GAG CCA GAT ATA GCC   480
98 Val Met Arg Thr Asn Asp Arg Phe Gly Arg Ala Glu Pro Asp Ile Ala
99 145                 150                 155                 160
101 GCG CTG GCG GAA CTG GCC GCG CTG CAG CTG CTC CCA CGT CTC AAT GAA   528
102 Ala Leu Ala Glu Leu Ala Ala Leu Gln Leu Leu Pro Arg Leu Asn Glu
103              165                 170                 175
105 GGC TTA GTG ATC ACC CAG GGA TTT ATC GGT AGC GAA AAT AAA GGT CGT   576
106 Gly Leu Val Ile Thr Gln Gly Phe Ile Gly Ser Glu Asn Lys Gly Arg
107              180                 185                 190
109 ACA ACG ACG CTT GGC CGT GGA GGC AGC GAT TAT ACG GCA GCC TTG CTG   624
110 Thr Thr Thr Leu Gly Arg Gly Gly Ser Asp Tyr Thr Ala Ala Leu Leu
111          195                 200                 205
113 GCG GAG GCT TTA CAC GCA TCT CGT GTT GAT ATC TGG ACC GAC GTC CCG   672
114 Ala Glu Ala Leu His Ala Ser Arg Val Asp Ile Trp Thr Asp Val Pro
115       210                 215                 220
117 GGC ATC TAC ACC ACC GAT CCA CGC GTA GTT TCC GCA GCA AAA CGC ATT   720
118 Gly Ile Tyr Thr Thr Asp Pro Arg Val Val Ser Ala Ala Lys Arg Ile
119 225                 230                 235                 240
121 GAT GAA ATC GCG TTT GCC GAA GCG GCA GAG ATG GCA ACT TTT GGT GCA   768
122 Asp Glu Ile Ala Phe Ala Glu Ala Ala Glu Met Ala Thr Phe Gly Ala
```

Input Set : A:\BB1037 US DIV CORRECTED SEQ LISTING.txt
Output Set: N:\CRF3\05102002\J023066A.raw

```
123              245              250              255    o
125 AAA GTA CTG CAT CCG GCA ACG TTG CTA CCC GCA GTA CGC AGC GAT ATC    816
126 Lys Val Leu His Pro Ala Thr Leu Leu Pro Ala Val Arg Ser Asp Ile
127              260              265              270
129 CCG GTC TTT GTC GGC TCC AGC AAA GAC CCA CGC GCA GGT GGT ACG CTG    864
130 Pro Val Phe Val Gly Ser Ser Lys Asp Pro Arg Ala Gly Gly Thr Leu
131              275              280              285
133 GTG TGC AAT AAA ACT GAA AAT CCG CCG CTG TTC CGC GCT CTG GCG CTT    912
134 Val Cys Asn Lys Thr Glu Asn Pro Pro Leu Phe Arg Ala Leu Ala Leu
135              290              295              300
137 CGT CGC AAT CAG ACT CTG CTC ACT TTG CAC AGC CTG AAT ATG CTG CAT    960
138 Arg Arg Asn Gln Thr Leu Leu Thr Leu His Ser Leu Asn Met Leu His
139 305              310              315              320
141 TCT CGC GGT TTC CTC GCG GAA GTT TTC GGC ATC CTC GCG CGG CAT AAT    1008
142 Ser Arg Gly Phe Leu Ala Glu Val Phe Gly Ile Leu Ala Arg His Asn
143              325              330              335
145 ATT TCG GTA GAC TTA ATC ACC ACG TCA GAA GTG AGC GTG GCA TTA ACC    1056
146 Ile Ser Val Asp Leu Ile Thr Thr Ser Glu Val Ser Val Ala Leu Thr
147              340              345              350
149 CTT GAT ACC ACC GGT TCA ACC TCC ACT GGC GAT ACG TTG CTG ACG CAA    1104
150 Leu Asp Thr Thr Gly Ser Thr Ser Thr Gly Asp Thr Leu Leu Thr Gln
151              355              360              365
153 TCT CTG CTG ATG GAG CTT TCC GCA CTG TGT CGG GTG GAG GTG GAA GAA    1152
154 Ser Leu Leu Met Glu Leu Ser Ala Leu Cys Arg Val Glu Val Glu Glu
155              370              375              380
157 GGT CTG GCG CTG GTC GCG TTG ATT GGC AAT GAC CTG TCA AAA GCC TGC    1200
158 Gly Leu Ala Leu Val Ala Leu Ile Gly Asn Asp Leu Ser Lys Ala Cys
159 385              390              395              400
161 GCC GTT GGC AAA GAG GTA TTC GGC GTA CTG GAA CCG TTC AAC ATT CGC    1248
162 Ala Val Gly Lys Glu Val Phe Gly Val Leu Glu Pro Phe Asn Ile Arg
163              405              410              415
165 ATG ATT TGT TAT GGC GCA TCC AGC CAT AAC CTG TGC TTC CTG GTG CCC    1296
166 Met Ile Cys Tyr Gly Ala Ser Ser His Asn Leu Cys Phe Leu Val Pro
167              420              425              430
169 GGC GAA GAT GCC GAG CAG GTG GTG CAA AAA CTG CAT AGT AAT TTG TTT    1344
170 Gly Glu Asp Ala Glu Gln Val Val Gln Lys Leu His Ser Asn Leu Phe
171              435              440              445
173 GAG TAA                                                           1350
174 Glu  *
W--> 175     450
177 (2) INFORMATION FOR SEQ ID NO: 2:
179      (i) SEQUENCE CHARACTERISTICS:
180           (A) LENGTH: 36 base pairs
181           (B) TYPE: nucleic acid
182           (C) STRANDEDNESS: single
183           (D) TOPOLOGY: linear
185      (ii) MOLECULE TYPE: DNA (genomic)
187      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
189 GATCCATGGC TGAAATTGTT GTCTCCAAAT TTGGCG                    36
```

```
191 (2) INFORMATION FOR SEQ ID NO: 3:
193        (i) SEQUENCE CHARACTERISTICS:
194            (A) LENGTH: 36 base pairs
195            (B) TYPE: nucleic acid
196            (C) STRANDEDNESS: single
197            (D) TOPOLOGY: linear
199       (ii) MOLECULE TYPE: DNA (genomic)
201       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
203 GTACCGCCAA ATTTGGAGAC AACAATTTCA GCCATG                   36
205 (2) INFORMATION FOR SEQ ID NO: 4:
207        (i) SEQUENCE CHARACTERISTICS:
208            (A) LENGTH: 48 base pairs
209            (B) TYPE: nucleic acid
210            (C) STRANDEDNESS: single
211            (D) TOPOLOGY: linear
213       (ii) MOLECULE TYPE: DNA (genomic)
215       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
217 CCCGGGCCAT GGCTACAGGT TTAACAGCTA AGACCGGAGT AGAGCACT   48
219 (2) INFORMATION FOR SEQ ID NO: 5:
221        (i) SEQUENCE CHARACTERISTICS:
222            (A) LENGTH: 37 base pairs
223            (B) TYPE: nucleic acid
224            (C) STRANDEDNESS: single
225            (D) TOPOLOGY: linear
227       (ii) MOLECULE TYPE: DNA (genomic)
229       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
231 GATATCGAAT TCTCATTATA GAACTCCAGC TTTTTTC                  37
233 (2) INFORMATION FOR SEQ ID NO: 6:
235        (i) SEQUENCE CHARACTERISTICS:
236            (A) LENGTH: 917 base pairs
237            (B) TYPE: nucleic acid
238            (C) STRANDEDNESS: single
239            (D) TOPOLOGY: linear
241       (ii) MOLECULE TYPE: DNA (genomic)
243       (ix) FEATURE:
244            (A) NAME/KEY: CDS
245            (B) LOCATION: 3..911
247       (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
249 CC ATG GCT ACA GGT TTA ACA GCT AAG ACC GGA GTA GAG CAC TTC GGC    47
250    Met Ala Thr Gly Leu Thr Ala Lys Thr Gly Val Glu His Phe Gly
251     1            5              10             15
253 ACC GTT GGA GTA GCA ATG GTT ACT CCA TTC ACG GAA TCC GGA GAC ATC   95
254 Thr Val Gly Val Ala Met Val Thr Pro Phe Thr Glu Ser Gly Asp Ile
255                20              25             30
257 GAT ATC GCT GCT GGC CGC GAA GTC GCG GCT TAT TTG GTT GAT AAG GGC  143
258 Asp Ile Ala Ala Gly Arg Glu Val Ala Ala Tyr Leu Val Asp Lys Gly
259               35              40             45
261 TTG GAT TCT TTG GTT CTC GCG GGC ACC ACT GGT GAA TCC CCA ACG ACA  191
262 Leu Asp Ser Leu Val Leu Ala Gly Thr Thr Gly Glu Ser Pro Thr Thr
```

Input Set : A:\BB1037 US DIV CORRECTED SEQ LISTING.txt
Output Set: N:\CRF3\05102002\J023066A.raw

```
263         50                    55                    60
265 ACC GCC GCT GAA AAA CTA GAA CTG CTC AAG GCC GTT CGT GAG GAA GTT   239
266 Thr Ala Ala Glu Lys Leu Glu Leu Leu Lys Ala Val Arg Glu Glu Val
267      65                    70                    75
269 GGG GAT CGG GCG AAG CTC ATC GCC GGT GTC GGA ACC AAC AAC ACG CGG   287
270 Gly Asp Arg Ala Lys Leu Ile Ala Gly Val Gly Thr Asn Asn Thr Arg
271  80                    85                    90                    95
273 ACA TCT GTG GAA CTT GCG GAA GCT GCT GCT TCT GCT GGC GCA GAC GGC   335
274 Thr Ser Val Glu Leu Ala Glu Ala Ala Ala Ser Ala Gly Ala Asp Gly
275                  100                   105                   110
277 CTT TTA GTT GTA ACT CCT TAT TAC TCC AAG CCG AGC CAA GAG GGA TTG   383
278 Leu Leu Val Val Thr Pro Tyr Tyr Ser Lys Pro Ser Gln Glu Gly Leu
279                  115                   120                   125
281 CTG GCG CAC TTC GGT GCA ATT GCT GCA GCA ACA GAG GTT CCA ATT TGT   431
282 Leu Ala His Phe Gly Ala Ile Ala Ala Ala Thr Glu Val Pro Ile Cys
283                  130                   135                   140
285 CTC TAT GAC ATT CCT GGT CGG TCA GGT ATT CCA ATT GAG TCT GAT ACC   479
286 Leu Tyr Asp Ile Pro Gly Arg Ser Gly Ile Pro Ile Glu Ser Asp Thr
287      145                   150                   155
289 ATG AGA CGC CTG AGT GAA TTA CCT ACG ATT TTG GCG GTC AAG GAC GCC   527
290 Met Arg Arg Leu Ser Glu Leu Pro Thr Ile Leu Ala Val Lys Asp Ala
291 160                   165                   170                   175
293 AAG GGT GAC CTC GTT GCA GCC ACG TCA TTG ATC AAA GAA ACG GGA CTT   575
294 Lys Gly Asp Leu Val Ala Ala Thr Ser Leu Ile Lys Glu Thr Gly Leu
295                  180                   185                   190
297 GCC TGG TAT TCA GGC GAT GAC CCA CTA AAC CTT GTT TGG CTT GCT TTG   623
298 Ala Trp Tyr Ser Gly Asp Asp Pro Leu Asn Leu Val Trp Leu Ala Leu
299                  195                   200                   205
301 GGC GGA TCA GGT TTC ATT TCC GTA ATT GGA CAT GCA GCC CCC ACA GCA   671
302 Gly Gly Ser Gly Phe Ile Ser Val Ile Gly His Ala Ala Pro Thr Ala
303                  210                   215                   220
305 TTA CGT GAG TTG TAC ACA AGC TTC GAG GAA GGC GAC CTC GTC CGT GCG   719
306 Leu Arg Glu Leu Tyr Thr Ser Phe Glu Glu Gly Asp Leu Val Arg Ala
307      225                   230                   235
309 CGG GAA ATC AAC GCC AAA CTA TCA CCG CTG GTA GCT GCC CAA GGT CGC   767
310 Arg Glu Ile Asn Ala Lys Leu Ser Pro Leu Val Ala Ala Gln Gly Arg
311 240                   245                   250                   255
313 TTG GGT GGA GTC AGC TTG GCA AAA GCT GCT CTG CGT CTG CAG GGC ATC   815
314 Leu Gly Gly Val Ser Leu Ala Lys Ala Ala Leu Arg Leu Gln Gly Ile
315                  260                   265                   270
317 AAC GTA GGA GAT CCT CGA CTT CCA ATT ATG GCT CCA AAT GAG CAG GAA   863
318 Asn Val Gly Asp Pro Arg Leu Pro Ile Met Ala Pro Asn Glu Gln Glu
319                  275                   280                   285
321 CTT GAG GCT CTC CGA GAA GAC ATG AAA AAA GCT GGA GTT CTA TAA TGAGAATTC   917
322 Leu Glu Ala Leu Arg Glu Asp Met Lys Lys Ala Gly Val Leu  *
323                  290                   295                   300
```

325 (2) INFORMATION FOR SEQ ID NO: 7:
327   (i) SEQUENCE CHARACTERISTICS:
328   (A) LENGTH: 22 base pairs

Input Set : A:\BB1037 US DIV CORRECTED SEQ LISTING.txt
Output Set: N:\CRF3\05102002\J023066A.raw

Please Note:
Use of n and/or Xaa have been detected in the Sequence Listing. Please review the
Sequence Listing to ensure that a corresponding explanation is presented in the <220>
to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:93; N Pos. 557,558,559,560,561
Seq#:102; N Pos. 97
Seq#:103; N Pos. 253,289,293
Seq#:104; Xaa Pos.32

Input Set : **A:\BB1037 US DIV CORRECTED SEQ LISTING.txt**
Output Set: **N:\CRF3\05102002\J023066A.raw**

L:31 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:32 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:175 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:775 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:29
L:842 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:31
L:1379 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:53
L:2635 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:104 after pos.:16